Authors' preprint version
The official version appears in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 4, pp. 2530-2540, April 2020.

PLADE: A Plane-based Descriptor for Point Cloud Registration with Small Overlap

Songlin Chen, Liangliang Nan, Renbo Xia, Jibin Zhao, and Peter Wonka

***Abstract*—Traditional point cloud registration methods require large overlap between scans, which imposes strict constraints on data acquisition. To facilitate registration, users have to carefully position scanners to ensure sufficient overlap. In this work, we propose to use high-level structural information (i.e., plane/line features and their inter-relationship) for registration, which is capable of registering point clouds with small overlap, allowing more freedom in data acquisition. We design a novel plane/line-based descriptor dedicated to establishing structure level correspondences between point clouds. Based on this descriptor, we propose a simple but effective registration algorithm[1]. We also provide a dataset[2] of real-world scenes containing a larger number of scans with a wide range of overlap. Experiments and comparisons with state-of-the-art methods on various datasets reveal that our method is superior to existing techniques. Though the proposed algorithm outperforms state-of-the-art methods on the most challenging dataset, the point cloud registration problem is still far from being solved, leaving significant room for improvement and future work.**

***Index Terms*—point cloud, registration, descriptor, scanning, dataset.**

I. Introduction

THE proliferation of acquisition devices (e.g., laser scanners and depth cameras) enables us to quickly obtain a massive volume of 3D point clouds of indoor and outdoor environments. The obtained point clouds have many applications in computer vision and computer graphics, including navigation and virtual/augmented reality. The nature of the scanning process typically results in a set of randomly oriented point clouds captured from different viewpoints, waiting to be registered. Although the registration problem has been extensively studied in the last decades, it still remains an open problem due to three main reasons.

Firstly, existing methods assume sufficient overlap between point clouds, which imposes restrictions on the scanning process, i.e., the user has to strategically position or move the scanner to ensure proper overlap between scans, making data acquisition a challenging task [1], [2]. In realistic scanning conditions, it is quite common that scans with insufficient overlap are obtained. This issue becomes vital when a scene is simultaneously scanned by multiple scanners and users. Another important scenario is when one wants to obtain complete scans of a scene, the user may apply a static laser scanner to capture the major part of the scene and a mobile scanner to complete the occluded regions. Scanning in such a fashion typically leads to a global point cloud and a set of local point clouds capturing local regions of the scene. These scans often have too small overlap for traditional registration methods to succeed.

Secondly, traditional registration methods focus on establishing correspondences between point clouds using local salient features. However, man-made scenes like building interiors and exteriors comprising mainly planar structures are common in the real-world [3], for which sufficient descriptive local features cannot be extracted for registration [4].

Thirdly, developing reliable point cloud registration approaches brings up significant challenges in evaluation tasks that involves capturing massive datasets and providing ground truth registrations. Unfortunately, very limited data sets are available and are typically created for specific environments (e.g., urban scenes) by using a single type of scanner (e.g., high-range laser scanners) [5], [6]) and typically have only a few scan pairs. The lack of diverse data sets (e.g., different environments, acquired using different sensors) and accurate ground-truth has caused various point cloud registration techniques to be poorly and unfairly evaluated [7], [8]. In fact, existing techniques can only be evaluated against small carefully crafted data sets.

In this work, we address the problem of registering point clouds with small overlap captured from real-world scenes. Since sufficient overlap and descriptive features cannot be guaranteed, our approach relies on high-level structures of the scene for registration. Specifically, man-made environments typically consist of planar structure, thus we represent the main structures of the scene as a collection of planes. These planar structures along with their inter-relations reveal high-level global characteristics of the scene and we believe that they provide sufficient information for registration. While there exists a fair amount of previous work using plane/line-based features, the robustness of existing plane-based methods is still not satisfactory [9], [10], [11], [12], [13], [14]. Our work proposes a plane/line-based descriptor to establish structure level correspondences between point clouds, with which robust registration can be effectively achieved.

In addition to the simple but effective registration algorithm, we provide a benchmark dataset scanned from a set of indoor and outdoor scenes with varying overlapping ratios, complementing existing datasets. As for evaluation, the performance of a registration method can be simply measured by the

S. Chen was with King Abdullah University of Science and Technology and Shenyang Institute of Automation, Chinese Academy of Sciences.
L. Nan is with the Deft University of Technology.
R. Xia and J. Zhao are with Shenyang Institute of Automation, Chinese Academy of Sciences.
P. Wonka is with King Abdullah University of Science and Technology.
Corresponding author: Liangliang Nan (liangliang.nan@gmail.com).

[1]The source code of PLADE is available at https://github.com/chsl/PLADE
[2]The dataset is available at https://3d.bk.tudelft.nl/liangliang/publications/2019/plade/resso.html

percentage of the successfully registered scans. Though experiments demonstrate that our method significantly outperforms the state of the art, a large portion of point clouds still remain unregistered. This indicates that the registration problem is far from being solved, allowing significant room for improvement. In summary, our main contributions include:

- a novel plane/line-based descriptor dedicated to establishing structure level correspondences between point clouds.
- a robust and fast point cloud registration algorithm using the plane/line-based descriptor, which significantly outperforms the state of the art.
- a benchmark dataset for evaluating point cloud registration algorithms. Our dataset contains scans with varying overlap, posing interesting challenges for research in point cloud registration.

II. Related Work

Point cloud registration methods can be roughly classified into two categories: coarse registration and fine registration. Fine registration algorithms aim to improve a given initial coarse registration. Such algorithms include ICP (Iterative Closest Point) [15] and its variants [16], [17], [18], [19]. In contrast, the inputs to the coarse registration algorithms are point clouds with unknown orientations. Thus, coarse registration is considered more challenging and has been receiving increasing attention in the past years. Our method falls into the coarse registration category. So in this section, we mainly discuss recent work on coarse registration, in particular, algorithms on local descriptor-based registration, global feature-based registration, and registration without overlap. For a comprehensive review of general point cloud registration algorithms, please refer to the survey by Maiseli et al. [8].

Local descriptor-based registration. Algorithms in this category are most popular in point cloud registration. These algorithms focus on using/defining local salient point features (i.e., transformation invariant descriptors) to establish point-wise correspondences between subsets (i.e., sets of key points) of the two point clouds [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30]. The typical procedure is to first extract key points and compute their descriptors and then establish sparse correspondences between the key points based on the descriptors. After that, various strategies have been developed to eliminate false correspondences. Commonly used techniques include geometric hashing [31] and RANSAC (Random Sample Consensus) [32], [33]. Other schemes are also developed for obtaining good correspondences. For example, Gelfand et al. [21] exploit a branch-and-bound algorithm to find the optimal set of correspondences. Based on the fact that certain ratios defined on a planar congruent set remain invariant under rigid transformations, Mellado et al. [24] proposes to extract all sets of coplanar 4-points to register point clouds with certain levels of noises and outliers. With initial correspondences computed using the Fast Point Feature Histogram (FPFH) feature [26], Zhou et al. propose an optimization framework that simultaneously suppresses spurious correspondences [34]. These methods demonstrated satisfactory performance on point clouds of general surfaces. However, they require sufficient overlap and are usually slow in processing large point clouds (e.g., scans of buildings).

Global feature-based registration. Compared to local features, global features cover larger scales of the point clouds and thus are more descriptive. The most widely used global feature is the plane feature that can be reliably and efficiently extracted from point clouds, especially for man-made scenes. These methods first segment the point clouds into planar patches and then search for correspondences at the patch level using various strategies [10], [11], [35], [12], [13].

Similar to local descriptors, various global shape descriptors have been developed for point cloud registration, such as the Hough Transform Descriptor, the Spherical Entropy Image [36], and the Viewpoint Descriptor [14]. By considering the layout of indoor scenes, Lee et al. propose to jointly estimate the layout and registration for indoor scene reconstruction [37].

Even higher-level features have also been studied in point cloud registration. Thapa et al. [38] propose a semantic feature-based method for registration of building scans. Their method starts with a semantic segmentation (achieved by using simple heuristics) of building scans. Then, correspondences are obtained by matching segments of the same semantic type and same pattern (topological relation with other features). Due to the difficulties in semantic segmentation, it remains unclear how to extend this method to register scans of general scenes/objects.

Registration without overlap. When overlap between scans is low, registration algorithms seek help from additional information provided by the point clouds [4], [39], [40]. Yan et al. propose to register building scans without overlap [4]. The inputs to their system are scans capturing multiple rooms of a large building, and/or scans capturing both the interior and exterior of a building. In their problem setting, the overlap between scans becomes extremely small or sometimes do not exist. The authors rely on portals (e.g., windows and doors) extracted from the point clouds to establish potential correspondences between scans. The global registration of the scans is then obtained by selecting a valid set of correspondences via a combinatorial optimization.

In recent years, researchers have also studied the problems of registering/assembling object pieces [41], [42], [43], [44]. Huang et al. [43] assemble fractured object pieces based on roughness analysis and patch-based features defined on fractured surfaces. Then, object pieces are registered by pairwise matching validated via penetration and consistency checks. Based on the fact that certain objects demonstrate continuous sharp feature curves, Huang et al. [44] align distinct object parts by enforcing the continuity of the sharp feature curves. This method relies heavily on the rich geometric features of the objects. Thus, it may not be scaled to practical scans of general scenes.

Our approach falls in global feature-based registration. We aim at registering featureless scans of real-world scenes that demonstrate unpredictable levels of overlap. We introduce a novel global descriptor that captures high-level structure information (i.e., inter-relationship of the major planes) of the scenes for registration.

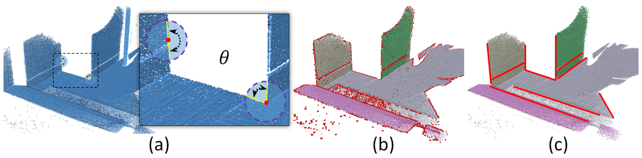


Fig. 1. Line extraction. (a) Input point cloud. (b) Extracted planar segments (in different colors) and boundary points (in red). (c) Extracted line segments (in red).

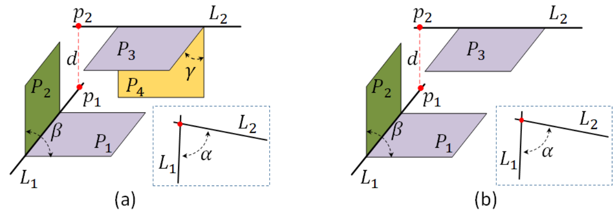


Fig. 2. The structure level descriptor. (a) A descriptor defined on two pairs of planes. Line L_1 is the intersection of planes P_1 and P_2; L_2 is the intersection of planes P_3 and P_4. (b) A descriptor defined on a line segment L_2 and two planes P_1 and P_2.

III. Our Approach

Our registration method is based on the traditional hypothesize-and-evaluate strategy. Specifically, the hypothesis is obtained by matching our novel plane/line-based descriptor, followed by removing the redundant matchings in the transformation space. Finally, the optimal registration is identified by evaluating the matching scores of the candidate registrations. In the following, we first describe our plane-based descriptor. We then detail our registration algorithm.

A. Structure Level Descriptor

A large portion of the traditional point cloud registration methods look into salient features and rely on a local geometric descriptor to establish correspondences between point clouds. Since sufficient overlaps and descriptive features may not be guaranteed, we rely on a high-level representation of the scene to achieve robust registration. We observe that man-made environments typically consist of planar structures, thus we represent the main structure of the scene as a collection of planes. These planar structures along with their inter-relations reveal high-level global characteristics of the scene, providing promising information for registration [9]. Specifically, we propose a structure level descriptor defined on planes and lines extracted from the point clouds, from which a unique rigid registration transformation can be established between two descriptors.

Plane/Line extraction. There exist a few approaches to extract basic geometric primitives (e.g., planes and lines) from point clouds [45], [46], [47]. As has been demonstrated that the RANSAC-based plane detection method is robust to noise and outliers and has been successfully applied to other tasks [48], we choose to utilize an efficient implementation of the RANSAC algorithm by Schnabel et al. [45] to extract planar segments from the point clouds. Figure 1 (b) shows an example of the extracted planar segments.

Given the planar segments, we then extract lines for each planar segment. Specifically, we first detect boundary points by looking into the distribution of the planar points within their neighborhood. We use an angle criterion to determine if a point is lying on the boundary of a planar segment. Figure 1 (a) inset illustrates our angle criterion (we choose the angle threshold θ to be $\frac{\pi}{2}$). Similar to plane extraction, we use a RANSAC strategy to extract line segments from the boundary points (see Figure 1 (c)). Alternative methods, such as [49], can also be applied to extract the lines segments directly from point clouds.

Defining the descriptor. Given a certain amount of planes abstracting the main structure of the scene, at least three non-parallel planes are required to establish rigid transformations between two point clouds. To avoid ambiguities (i.e., a corner of three planes can be matched to multiple similar corners) and obtain a unique transformation, we look into quadruplets of non-parallel planes.

We first compute pairwise intersections of the supporting planes of the extracted planar segments, resulting in a set of lines. To cope with near co-planar planes, we discard a line L_i if $dist(L_i, c) > r$, where c and r denote the center and radius of the bounding sphere of the point cloud. Figure 2 (a) illustrates the primitives (i.e., four planes) on which our plane-based descriptor is defined. Specifically, the plane-based descriptor is an eight-dimensional vector consisting of the following entries

- d: the distance between the two lines L_1 and L_2;
- $\angle(L_1, L_2)$: the angle between L_1 and L_2;
- $\angle(P_1, P_2)$ and $\angle(P_3, P_4)$: angles introduced by the two pairs of planes;
- $\angle(L_1, P_3)$, $\angle(L_1, P_4)$, $\angle(L_2, P_1)$, and $\angle(L_2, P_2)$: the angles between the intersecting lines of two planes and the other planes;

Note that we choose the acute angle for each pair of primitives. To ensure descriptiveness, our plane-based registration descriptor is defined depending on the relative magnitudes of the angles between primitives

$$\mathbf{d^8} = \begin{bmatrix} dist(L_1, L_2) \\ \angle(L_1, L_2) \\ \angle(P_1, P_2) \\ \angle(P_3, P_4) \\ \min(\angle(L_1, P_3), \angle(L_1, P_4)) \\ \max(\angle(L_1, P_3), \angle(L_1, P_4)) \\ \min(\angle(L_2, P_1), \angle(L_2, P_2)) \\ \max(\angle(L_2, P_1), \angle(L_2, P_2)) \end{bmatrix} \tag{1}$$

if $\angle(P_1, P_2) < \angle(P_3, P_4)$. Otherwise, we change the order of the planes and then define the descriptor. Here $\min(*, *)$ and $\max(*, *)$ indicate the smaller and greater value of two angles, respectively.

The above plane-based registration descriptor is defined purely on two pairs of non-parallel planes, with which a unique rigid registration transformation can be established between two descriptors. In the very unlikely cases (in particular when the overlap between the point cloud pair is small), less than two pairs of non-parallel planes can be found. The point cloud shown in Figure 1 is such an example, where only two parallel horizontal planes and two parallel vertical planes are extracted. Thus, no quadruplet of non-parallel planes exists to uniquely define a rigid transformation. In such a case, we seek help from additional line features of the scene. So in addition to the eight-dimensional plane-based descriptor, we also define another type of registration descriptor on a smaller number of geometric primitives, i.e., a pair of non-parallel planes and a line segment. Similarly, the plane/line-based descriptor is a six-dimensional vector defined as

$$\mathbf{d^6} = \begin{bmatrix} dist(L_1, L_2) \\ \angle(L_1, L_2) \\ \angle(P_1, P_2) \\ \angle(L_1, P_3) \\ \min(\angle(L_2, P_1), \angle(L_2, P_2)) \\ \max(\angle(L_2, P_1), \angle(L_2, P_2)) \end{bmatrix} \quad (2)$$

B. Registration

With the structure level registration descriptor, we are now able to compute transformations between two point clouds. Since our descriptor characterize the inter-relation between non-parallel planes/lines, we can establish a unique rigid transformation using a descriptor $\mathbf{d}_{\mathcal{L}}$ from a point cloud $\mathcal{L}$ and its best-matched descriptor $\mathbf{d}_{\mathcal{G}}$ from the other point cloud $\mathcal{G}$.

We enumerate all plane/line combinations to collect a set of descriptors in both $\mathcal{L}$ and $\mathcal{G}$, namely $\mathbf{D}_{\mathcal{G}} = \mathbf{D}^8_{\mathcal{G}} \cup \mathbf{D}^6_{\mathcal{G}}$ and $\mathbf{D}_{\mathcal{L}} = \mathbf{D}^8_{\mathcal{L}} \cup \mathbf{D}^6_{\mathcal{L}}$, where $\mathbf{D}^8_* = \{\mathbf{d}^8_*\}$ and $\mathbf{D}^6_* = \{\mathbf{d}^6_*\}$ denote the eight dimensional and six dimensional descriptors respectively.

Descriptor matching. To efficiently find the best matches of descriptor pairs, we build a KD-tree for the descriptors $\mathbf{D}_{\mathcal{G}}$ and we query the most similar descriptor for each descriptor in $\mathbf{D}_{\mathcal{L}}$. The distance of a descriptor pair is computed as the Euclidean distance of the two descriptors. To compare a 6D descriptor against an 8D descriptor, we simply exclude the two extra dimensions from the 8D descriptor vector. By doing this, the 8D descriptor vector is degraded to 6D. So the Euclidean distance between them can be computed using the corresponding entries.

Transformation redundancy. Since our registration descriptor mainly encodes the geometric information of the planes, simply enumerating all combinations of the planes results in duplicated transformations. Figure 3 visualizes the computed translations and rotations from the best-matched descriptor pairs of two point clouds. We can see that a large portion of the transformations is duplicated. This can be observed from the large number of points but fewer clusters in the visualization.

Given a large number of transformations computed from the best-matched descriptor pairs, our final goal is to choose the best transformation that can register the two point clouds. To achieve this goal, we have to evaluate the confidence for each transformation. Here, the confidence of the transformation is typically measured by the number of matched points. Precisely measuring the number of matched points requires querying the nearest neighbor for every point in one point cloud. Performing such queries on small numbers of transformations is affordable. However, the large portion of duplicated transformations hinders us from efficiently obtaining the optimal transformation. To this end, we first remove the redundancy in the transformations and we keep only the most representative ones. Using a KD-tree structure, we search for the neighbors $\mathbf{T}_i$ of each transformation $\mathbf{t}_i$ within a radius r. We simply replace $\mathbf{t}_i \cup \mathbf{T}_i$ with their mass center. In our implementation, we chose $r_t = 0.001 \cdot r_{\mathcal{L}}$ for translations and $r_r = 2°$ for rotations, where $r_{\mathcal{L}}$ denotes the radius of the bounding sphere of the point cloud $\mathcal{L}$. After the redundancy being removed, the number of transformations is significantly reduced. Then, we perform penetration tests to further reduce infeasible transformations. To do so, we look into the point distribution of two planar segments (see Figure 4). Penetration is considered occurring only if the points of each planar segment lie on both sides of the supporting plane of the other planar segment.

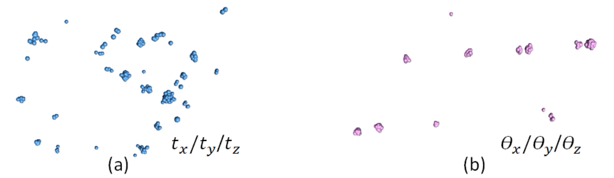


Fig. 3. A visualization of the computed translations and rotations from the best-matched descriptor pairs of two point clouds. Each point in (a) represents a translation and each point in (b) represents a rotation (denoted by the three angles w.r.t. the axes). Note that minor jittering has been added to reveal the duplicated transformations.

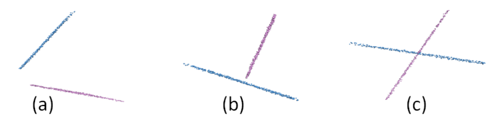


Fig. 4. Penetration tests for two planar segments. (a) and (b) do not have penetration. (c) An example of penetration.

Identifying optimal registration. Intuitively, the optimal registration transforms the point cloud $\mathcal{L}$ in a way such that the most number of points can be matched to the points in the point cloud $\mathcal{G}$. This is true for most cases, especially for objects with curved surfaces. However, when dealing with man-made scenes that typically comprise planar regions, the transformation receiving most matched points does not always suggest the optimal registration. This is obvious because one planar segment (a set of points lying in a plane) can be matched with any other planar segments. In this work, we measure the confidence of a registration transformation (i.e., a translation denoted by $\mathbf{t}$ and a rotation denoted by $\mathbf{r}$) by combining two criteria

$$conf(\mathbf{t}, \mathbf{r}) = w_{plane} \cdot R_{plane} + w_{points} \cdot R_{points}, \quad (3)$$

where R_{plane} and R_{points} denote the ratio of the matched planes and the ratio of the matched points respectively. The two weights w_{plane} and w_{points} are empirically chosen to be 0.2 and 0.8, respectively. By computing the registration confidence for each transformation, the one with the highest registration confidence is considered as the optimal coarse registration.

IV. BENCHMARK DATASET

To evaluate our method and, more importantly, to provide a more practical benchmark dataset complementing existing datasets, we create a new dataset **RESSO**[2] targeting both indoor and outdoor scenes.

Data collection. Our data acquisition involves two different types of commercial scanners: a high-range static laser scanner (Leica ScanStation C10, with an effective operating range of 100 meters) and a hand-held scanner (FARO Freestyle X, operating range 3 meters). These two scanners have significantly different operating ranges, accuracy, and resolutions, posing sufficient challenges to registration algorithms.

We scanned 187 point clouds in total for 15 different environments (10 indoor scenes and 5 outdoor scenes). Each indoor scene is captured by a few global point clouds using the static laser scanner and optionally multiple local point clouds using the hand-held scanner. The global point clouds capture the majority of each indoor scene and the local point clouds are intended to capture local regions of the scene, especially the regions that are occluded in the global point clouds. This further adds to the challenges for registration. Due to larger sizes, the outdoor scenes are mainly captured using the long-range static laser scanner.

We scanned the scenes without adding additional clutter for augmenting naturally occurring features and we tried to create some overlap, but not an excessive amount. Also, the fact that the point clouds stem from different scanners is a possible challenge for some feature extractors.

Overlap between scans. Real-world scans typically have unpredictable varying overlap ratios, which is challenging to registration algorithms. We choose to quantify the overlap of two point clouds by measuring the percentage of points that have the closest corresponding point (in another scan of the pair) closer than a threshold ϵ. Considering noises in the input point clouds and the unavoidable errors in the registration, we compute the ϵ-overlap for each scan pair at a discrete set of ϵ values. We depict these discrete ϵ-overlap values in a curve, so as to intuitively reveal the overlap between scans. Figure 5 demonstrates the ϵ-overlap curves for a few point cloud pairs from RESSO and other datasets. From the ϵ-curves, we can see that RESSO has less but a wider range of overlap. Thus, our new dataset is a more challenging and a useful complement to existing datasets.

Ground truth registration. Given the challenges in the registration problem itself and the large number of scans, we obtain ground truth registrations using a combination of automatic approaches and manual registration. Specifically, we run our registration algorithm on the point clouds of each scene and we record the transformation matrices of the successfully registered point clouds by visual inspection and fine tuning of the registration using ICP [15]. For those failed to be registered in the automatic phase, we manually registered them as initialization to ICP.

[2]**RESSO**: **Re**al-world **S**cans with **S**mall **O**verlap.

V. RESULTS AND DISCUSSION

We implemented our method in C++ using the Point Cloud Library [51]. In our current implementation, we mainly focus on local registration (i.e., pairs of the scans), leaving global registration (i.e., simultaneously registering all scans in a scene) as future work. Experiments on various datasets demonstrated that our method significantly outperforms state-of-the-art registration techniques.

Evaluation method. Our work focuses on coarse registration, but in practice, fine registration might be used as a post-process. One possible evaluation method would be to evaluate the combination of coarse and fine registration algorithms. We opt for a more direct evaluation, where we separately evaluate the impact of coarse registration and fine registration results by comparing the transformed scans to their ground truth. While there are many fine registration methods, we use ICP [15] as a popular representative. Specifically, we consider a registration successful if the registered scan is close enough to the ground truth, i.e.,

$$dist(\mathbf{s}_r, \mathbf{s}_g) > d_t, \tag{4}$$

where $dist(\mathbf{s}_r, \mathbf{s}_g)$ measures the average point distance between a registered scan $\mathbf{s}_r$ and the ground truth $\mathbf{s}_g$. To choose an appropriate value for the threshold d_t, we take into consideration that our coarse registration result is provided as initialization to a fine registration method. We conducted multiple experiments and we present the one on all the point cloud pairs in Figure 9 here. We introduced a sequence (i.e., 10) of random perturbation transformations (starting from the ground truth transformation) such that the mean distance of all the corresponding points was increased at a constant interval of $5cm$. Then we ran the ICP algorithm of [15] on all the point cloud pairs in each sequence to test if ICP could converge. We recorded the success rate for each sequence and the result is demonstrated in Table I. This experiment showed that ICP converged when the mean distance was smaller than $20cm$ for indoor scenes and $25cm$ for outdoor scenes. Based on these experiments, we conservatively set d_t to $10cm$ for indoor scenes and $20cm$ for outdoor scenes.

Registration results. Figures 6 and 7 visualize the registration results of the proposed method on ten indoor scenes and five outdoor scenes from our dataset RESSO, respectively. Thanks to the descriptive plane-based descriptor, our registration method managed to register all these scan pairs. Though the indoor scene in Figure 6 (j) and the outdoor scene in Figure 7 (d) partially consist of curved surfaces, planar structures still dominate and our method successfully registered these point clouds. The outdoor scene shown in Figure 7 (a) contains many trees. The planar regions still provide sufficient information for a reliable registration. Besides RESSO, we also tested our registration method on point clouds from publicly available datasets and related works. The visual results are shown in Figure 8.

Our method is capable of registering scans with small overlap. Figures 6, 7, and 8 show the registration results of all point clouds for each scene, thus it is difficult to observe the overlaps between scans. In Figure 9, we demonstrate a few pairs of scans from our results shown in Figures 6 and 7.

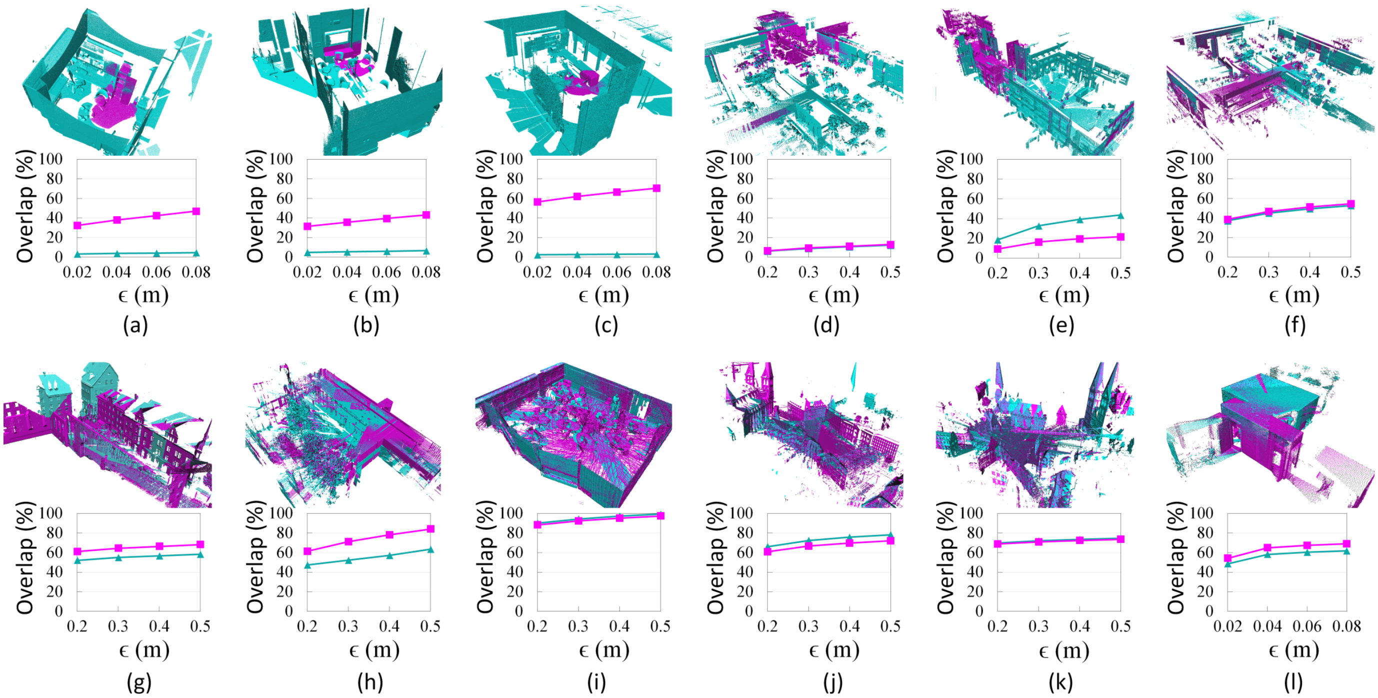


Fig. 5. Overlap of point cloud pairs. Top row: scan pairs from RESSO. Bottom row: scan pairs from existing datasets, i.e., (g) ETH [50], (h) DS2-L [13], (i) TLS-ZEB [14], (j) and (k) Robotic 3D scan repository [5], (l) DS1-H [6]. The corresponding ϵ-overlap curve is shown below each scan pair.

TABLE I

SUCCESS RATE (%) OF ICP [15] ON A SEQUENCE OF POINT CLOUD PAIRS WITH INCREASING PERTURBATION LEVELS.

Figure \ Perturbation (cm)	5	10	15	20	25	30	35	40	45
Figure 9 (a)	100	100	100	100	40	0	0	0	0
Figure 9 (b)	100	100	100	100	70	0	0	0	0
Figure 9 (c)	100	100	100	100	100	30	0	0	0
Figure 9 (d)	100	100	100	100	100	100	90	30	0

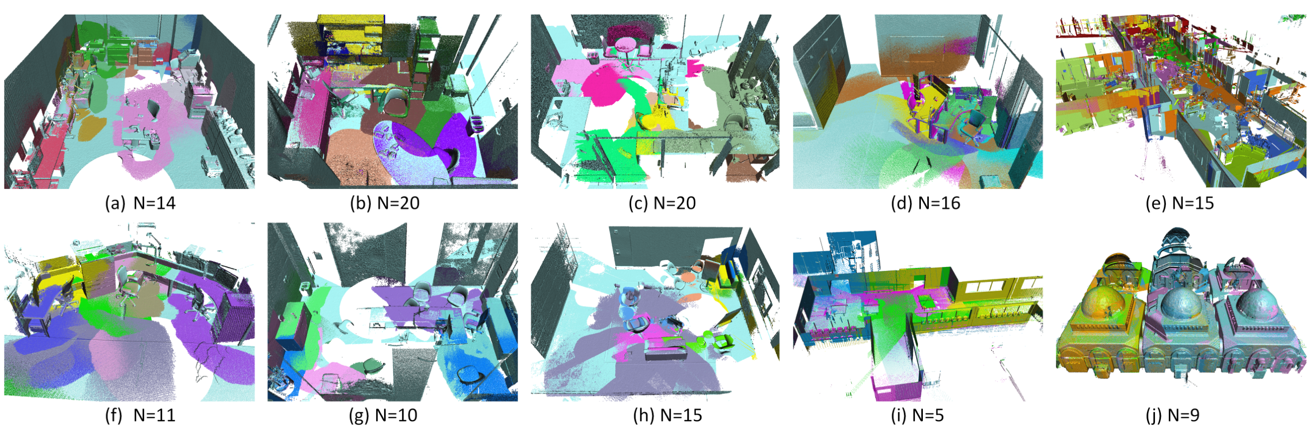


Fig. 6. Registration results of the indoor scenes from RESSO. The ceilings have been removed to better reveal the building interiors. The number bellow each subfigure indicates the total scans in each scene.

Initialization to fine registration. To test if our coarse registration results can be further improved by a fine registration method, we ran the ICP algorithm of [15] on all the point cloud pairs shown in Figure 9 and recorded the registration error before and after the ICP step. The result is reported in Table II. We can see that the ICP step significantly reduced the registration error compared to the that of the coarse registration, indicating that our coarse registration results provided good initialization to the ICP algorithm.

Robustness to plane detection. Our plane-based descriptor is designed to capture the global structure of a scene, allowing us to reliably establish structure level correspondences

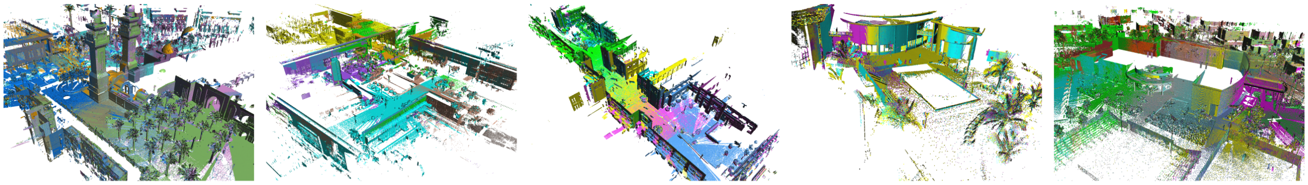


Fig. 7. Registration results of the outdoor scenes from RESSO. The number bellow each subfigure indicates the total scans in each scene.

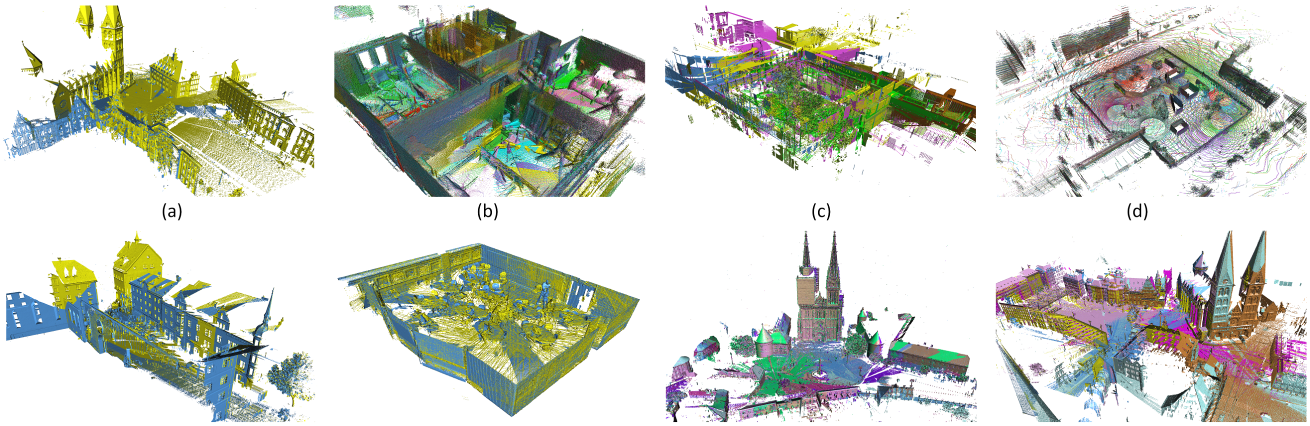


Fig. 8. Registration results of our method on various datasets. (a) Bremen [52], (b) DS1-H [6], (c) DS2-L [13], (d) DS3-V [13], (e) ETH [50], (f) TLS-ZEB [14], (g) and (h) Robotic 3D scan repository [5]. The ceilings in (b), (c), and (f) have been removed to better reveal the building interiors.

TABLE II
REGISTRATION ERRORS BEFORE AND AFTER APPLYING THE FINE REGISTRATION METHOD OF [15] ON THE CLOUD PAIRS SHOWN IN FIGURE 9.

Figure \ Errors (m)	Coarse	Fine
Figure 9 (a)	0.026	0.008
Figure 9 (b)	0.025	0.006
Figure 9 (c)	0.154	0.039
Figure 9 (d)	0.136	0.042

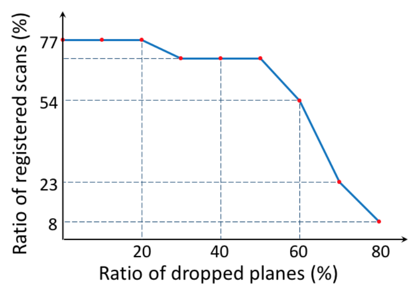


Fig. 10. Registration by gradually dropping planes on the scene show in Figure 6 (a). Planes with smaller numbers of points are dropped first.

between two point clouds. Since a few descriptive planes are adequate in depicting the main structure of the scene, it is not necessary (nor possible) to obtain a complete set of planes accurately extracted from the point clouds. To evaluate this, we repeatedly ran our method on the scene shown in Figure 6 (a) by incrementally removing planes. Specifically, we remove 10% of the extracted planes at each iteration until our algorithm breaks down. Figure 10 reports how our method behaves by gradually dropping planes. Such a test confirms that a few dominant planes can provide adequate information for point cloud registration, allowing our method to achieve satisfactory registration results as long as certain descriptive planes (i.e., a small portion of planes) are present.

Robustness to noise. In order to evaluate the impact of noisy surfaces, we added Gaussian noise to a pair of point clouds from Figure 6 (a) with increasing noise levels, i.e., standard deviations (σ) $15cm$, $30cm$, $45cm$, and $60cm$, respectively. Though the noise levels are quite high, we were still able to extract planes with sufficient quality at three noise levels. Figure 11 shows the registration result at noise level $\sigma = 45cm$. However, when the noise level reached $60cm$, where the smaller point cloud (in green) were completely contaminated by the noise (note how difficult to recognize the chairs in the scene), our RANSAC-based plane extraction algorithm failed to detect sufficient planes to establish reliable correspondences for the registration. Such a test indicates

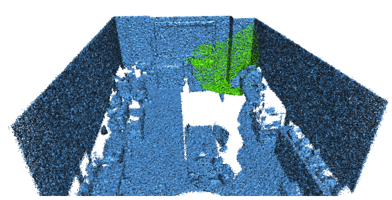


Fig. 11. Registration of two point clouds with Gaussian noise (standard deviation $\sigma = 45cm$).

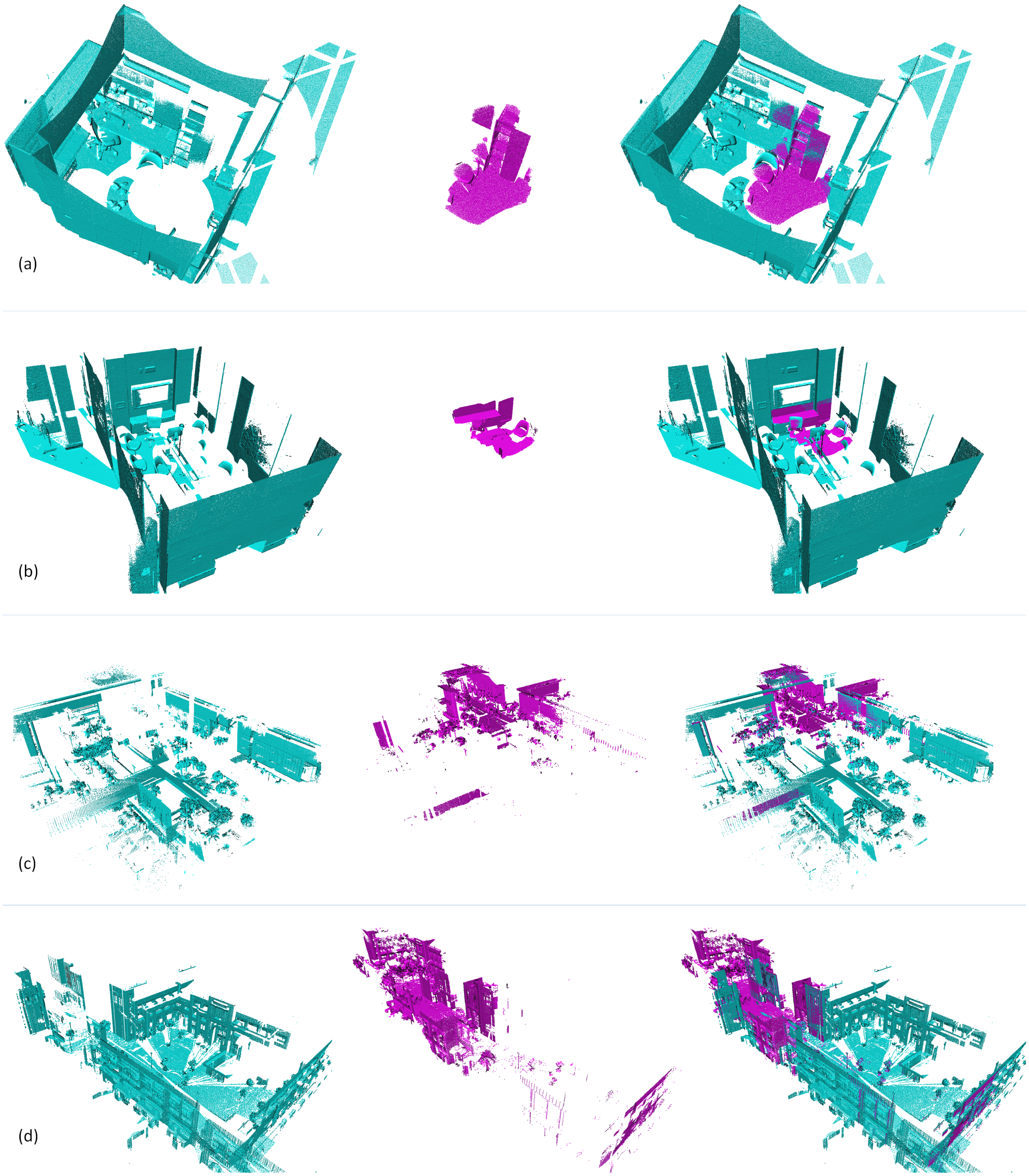


Fig. 9. Pairs of point clouds registered by our method, intended to reveal the overlaps between scans. The first two columns show the input scan pairs and the right column shows the registration results.

that our method is robust to noise as long as the major representative planes can be extracted.

Comparison. We compared our method against various point cloud registration methods, including local descriptor-based approaches and plane-based approaches. Table III and Table IV report the performance of our method and the competing methods on some of scan pairs from scenes shown in Figures 6 and 7. The performance is measured in terms of the percentage of successfully registered point clouds. From Table III, we can see that Super4PCS [24] failed in registering most of the point cloud pairs from the indoor scenes. Other local descriptor-based registration methods managed to register only a small portion of the scans. Such poor performance is mainly due to the small overlap and the absence of local geometric features in the point clouds. As expected, the performance of these techniques improves when the scans have significantly larger overlap, e.g., the indoor scene shown in Figure 6 (j) and the outdoor scenes shown in Figure 7. Besides, the scenes in Figure 6 (j) and Figure 7 (a) contain some curved structures, adding descriptive local geometric features for registration. The large overlapping ratio and the geometric features bring the performance improvements.

We also compared our method against various state-of-the-art plane-based registration methods. Due to that source code of the completing methods is not available, we ask the authors to run their algorithms on a few scan pairs randomly chosen from RESSO. These scan pairs demonstrate a wide range of overlapping ratios. Table IV summarizes the comparison. Among these methods, the RANSAC-based approach is quite similar to our method, except that we replace our descriptor-based correspondence search with RANSAC-based correspondence search. From all these comparisons, we can conclude that planes are effective features for registering scans of real-world scenes. Based on the novel plane-based , our method significantly outperforms the competing methods in terms of the percentage of successfully registered scans.

Running times. Table V gives the running times of our method and the competing methods on the scenes shown in Figures 6 and 7. The Super4PCS algorithm [24] requires to explore sufficient sets of coplanar 4-points and thus becomes more expensive for scans of large scenes. The method by Zhou et al. [34] demonstrate higher efficiency than the Super4PCS technique because its optimization process involves neither correspondence updates nor closest-point queries. Compared to these techniques, our method takes advantage of the plane-based descriptor so that structure level correspondences between scans can be very efficiently established via nearest neighbor search. Thus, it has better efficiency than most of the competing methods, in particular for larger scenes. Note that the input point clouds were down-sampled to enable the competing algorithms to generate their results within an acceptable time frame.

Limitations. Our plane-based descriptor is dedicated to registering point clouds of scenes that at least partially consist of planar structures. Thus, the proposed descriptor is especially suitable for registering scans of man-made environments. The descriptor will probably not be successful for scans of vegetation and scans of individual objects that consist of curved surfaces.

Another limitation of our current implementation is that the confidence metric defined in Equation 3 can only handle the majority of the tested point clouds in our benchmark data sets. It still remains a challenge to develop a reliable confidence metric that works for all scenarios.

VI. Conclusions

In this paper, we discussed several challenges of the point cloud registration problem. To address these challenges, we presented a simple but effective method for registering practical and feature-poor scans with small overlap in arbitrary initial poses. Our method is based on a high-level descriptor that reveal structural characteristics of the scenes, leading to superior registration performance.

Despite the excellent performance of the proposed registration algorithm, we demonstrated that the point cloud registration problem is far from being solved, leaving significant room for improvement and future work. We also provide the community a new challenging benchmark dataset that is large and challenging enough to ensure that registration algorithms are fairly evaluated and compared, in the hope that experts in related fields seize such research opportunities and push the state of the art in point cloud registration forward.

Acknowledgements

We would like to thank Yu Tian for helping us with data acquisition. This research was supported by the KAUST Office of Sponsored Research (award No. OCRF-2014-CGR3-62140401) and the Visual Computing Center (VCC) at KAUST. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for rendering the point clouds in this research.

References

[1] W. R. Scott, G. Roth, and J.-F. Rivest, “View planning for automated three-dimensional object reconstruction and inspection,” *ACM Comput. Surv.*, vol. 35, no. 1, pp. 64–96, 2003.

[2] S. Wu, W. Sun, P. Long, H. Huang, D. Cohen-Or, M. Gong, O. Deussen, and B. Chen, “Quality-driven poisson-guided autoscanning,” *ACM Trans. Graph.*, vol. 33, no. 6, pp. 203:1–203:12, 2014.

[3] A. Monszpart, N. Mellado, G. J. Brostow, and N. J. Mitra, “Rapter: Rebuilding man-made scenes with regular arrangements of planes,” *ACM Trans. Graph.*, vol. 34, no. 4, pp. 103:1–103:12, 2015.

[4] F. Yan, L. Nan, and P. Wonka, “Block assembly for global registration of building scans,” *SIGGRAPH Asia 2016*, vol. 35, no. 6, pp. 237:1–237:11.

[5] A. Nüchter and K. Lingemann, “Robotic 3d scan repository,” *Jacobs University Bremen gGmbH and University of Osnabrück*, 2011.

[6] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart, “Challenging data sets for point cloud registration algorithms,” *The International Journal of Robotics Research*, vol. 31, no. 14, pp. 1705–1711, 2012.

[7] M. Magnusson, N. Vaskevicius, T. Stoyanov, K. Pathak, and A. Birk, “Beyond points: Evaluating recent 3d scan-matching algorithms,” in *ICRA 2015*, 2015, pp. 3631–3637.

[8] B. Maiseli, Y. Gu, and H. Gao, “Recent developments and trends in point set registration methods,” *Journal of Visual Communication and Image Representation*, 2017.

[9] C. Brenner, C. Dold, and N. Ripperda, “Coarse orientation of terrestrial laser scans in urban environments,” *ISPRS journal of photogrammetry and remote sensing*, vol. 63, no. 1, pp. 4–18, 2008.

[10] K. Khoshelham, “Automated localization of a laser scanner in indoor environments using planar objects,” in *2010 International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 2010, pp. 1–7.

[11] F. Bosché, “Plane-based registration of construction laser scans with 3d/4d building models,” *Advanced Engineering Informatics*, vol. 26, no. 1, pp. 90–102, 2012.

[12] Y. Xu, R. Boerner, W. Yao, L. Hoegner, and U. Stilla, “Automated coarse registration of point clouds in 3d urban scenes using voxel based plane constraint.” *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 4, 2017.

[13] J. Sanchez, F. Denis, P. Checchin, F. Dupont, and L. Trassoudaine, “Global registration of 3d lidar point clouds based on scene features: Application to structured environments,” *Remote Sensing*, vol. 9, no. 10, p. 1014, 2017.

[14] W. Forstner and K. Khoshelham, “Efficient and accurate registration of point clouds with plane to plane correspondences,” in *ICCV 2017*.

[15] P. J. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *IEEE TPAMI*, vol. 14, no. 2, pp. 239–256, 1992.

[16] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, “Nonrigid registration using free-form deformations: application to breast mr images,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

TABLE III
COMPARISON WITH A FEW LOCAL DESCRIPTOR-BASED METHODS AND SUPER4PCS [24]. THE PERFORMANCE HERE IS MEASURED BY THE PERCENTAGE (%) OF THE SCANS THAT WERE SUCCESSFULLY REGISTERED IN EACH SCENE.

Method \ **Figure**	6 (a)	6 (b)	6 (c)	6 (d)	6 (f)	6 (g)	6 (h)	6 (j)	7 (a)	7 (b)	7 (c)
Super4PCS [24]	0	0	5	0	0	0	7	13	9	33	20
FPFH [25]	15	11	5	7	0	11	14	33	42	17	40
FastGlobal [34]	31	21	21	47	20	22	7	33	92	50	100
CZK [53]	46	37	21	53	50	22	50	22	75	33	100
RANSAC-based	15	16	16	13	50	67	43	0	43	33	100
Ours	**77**	**53**	**37**	**67**	**70**	**78**	**64**	**50**	**100**	**100**	**100**

TABLE IV
COMPARISON WITH A FEW PLANE-BASED REGISTRATION METHODS ON RESSO. THE RATE IS MEASURED BY THE PERCENTAGE (%) OF THE TOTAL REGISTRATIONS EACH METHOD SUCCEEDED.

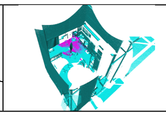
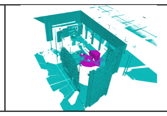
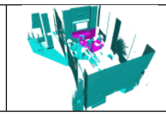
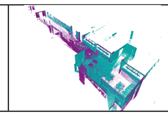
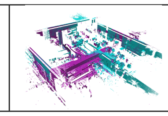
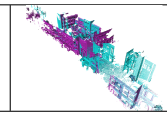
Method \ **Data**							**Rate (%)**
VBPC [12]				✓	✓	✓	50
SSFR [13]	✓		✓	✓	✓	✓	83
RANSAC-based	✓			✓		✓	50
Ours	✓	✓	✓	✓	✓	✓	100

TABLE V
EXECUTION TIMES (IN SECONDS) OF OUR METHOD AND SOME COMPETING METHODS ON THE SCENES SHOWN IN FIGURES 6 AND 7. THE EXECUTION TIMES WERE MEASURED ON A LAPTOP WITH A DUAL-CORE 2.4 GHZ INTEL CORE I3-4000M CPU.

Method \ **Figure**	6 (a)	6 (b)	6 (c)	6 (d)	6 (f)	6 (g)	6 (h)	6 (j)	7 (a)	7 (b)	7 (c)
Super4PCS [24]	202	305	514	349	631	434	534	527	479	521	759
FPFH [25]	198	20	61	130	107	51	79	84	477	260	270
FastGlobal [34]	10	2	3	7	11	4	7	21	119	59	31
CZK [53]	17	7	9	8	16	9	11	59	60	63	57
RANSAC-based	504	223	587	542	168	450	98	3107	2478	5439	3196
Ours	25	6	8	13	13	9	9	5	21	23	16

[17] S. Rusinkiewicz and M. Levoy, “Efficient variants of the icp algorithm,” in *3-D Digital Imaging and Modeling*. IEEE, 2001, pp. 145–152.

[18] H. Li, R. W. Sumner, and M. Pauly, “Global correspondence optimization for non-rigid registration of depth scans,” in *Computer graphics forum*, vol. 27, no. 5, 2008, pp. 1421–1430.

[19] A. Segal, D. Haehnel, and S. Thrun, “Generalized-icp.” in *Robotics: Science and Systems*, vol. 2, no. 4, 2009.

[20] A. E. Johnson, “Spin-images: a representation for 3-d surface matching,” Ph.D. dissertation, 1997.

[21] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, “Robust global registration,” in *SGP 2005*, vol. 2, no. 3, p. 5.

[22] X. Li and I. Guskov, “Multi-scale features for approximate alignment of point-based surfaces,” in *SGP 2005*.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[24] N. Mellado, D. Aiger, and N. J. Mitra, “Super 4pcs fast global pointcloud registration via smart indexing,” *Computer Graphics Forum*, vol. 33, no. 5, pp. 205–215, 2014.

[25] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, “Aligning point cloud views using persistent feature histograms,” in *IROS 2008*, pp. 3384–3391.

[26] R. B. Rusu, N. Blodow, and M. Beetz, “Fast point feature histograms (fpfh) for 3d registration,” in *ICRA 2009*, 2009, pp. 3212–3217.

[27] J. Yang, Z. Cao, and Q. Zhang, “A fast and robust local descriptor for 3d point cloud registration,” *Inf. Sci.*, vol. 346, no. C, pp. 163–179, 2016.

[28] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, “3dmatch: Learning local geometric descriptors from rgb-d reconstructions,” in *CVPR 2017*.

[29] X. Huang, “Learning a 3d descriptor for cross-source point cloud registration from synthetic data,” *arXiv preprint arXiv:1708.08997*, 2017.

[30] G. Elbaz, T. Avraham, and A. Fischer, “3d point cloud registration for localization using a deep neural network auto-encoder,” in *CVPR 2017*.

[31] H. J. Wolfson and I. Rigoutsos, “Geometric hashing: An overview,” *IEEE Comput. Sci. Eng.*, vol. 4, no. 4, pp. 10–21, 1997.

[32] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[33] C.-S. Chen, Y.-P. Hung, and J.-B. Cheng, “Ransac-based darces: A new approach to fast automatic registration of partially overlapping range images,” *IEEE TPAMI*, vol. 21, no. 11, pp. 1229–1234, 1999.

[34] Q.-Y. Zhou, J. Park, and V. Koltun, “Fast global registration,” in *ECCV 2016*, 2016, pp. 766–782.

[35] J. Xiao, B. Adler, J. Zhang, and H. Zhang, “Planar segment based three-dimensional point cloud registration in outdoor environments,” *Journal of Field Robotics*, vol. 30, no. 4, pp. 552–582, 2013.

[36] B. Sun, “3d global shape descriptors applied in scan registration,” Ph.D. dissertation, Hamburg, Universität Hamburg, Diss., 2015.

[37] J.-K. Lee, J.-W. Yea, M.-G. Park, and K.-J. Yoon, “Joint layout estimation and global multi-view registration for indoor reconstruction,” in *ICCV 2017*.

[38] A. Thapa, S. Pu, and M. Gerke, “Semantic feature based registration of terrestrial point clouds.” Citeseer.

[39] A. Cohen, J. L. Schönberger, P. Speciale, T. Sattler, J.-M. Frahm, and M. Pollefeys, “Indoor-outdoor 3d reconstruction alignment,” in *European Conference on Computer Vision*. Springer, 2016, pp. 285–300.

[40] T. Koch, M. Körner, and F. Fraundorfer, “Automatic alignment of indoor

and outdoor building models using 3d line segments," *Proceedings of Computer Vision and Pattern Recognition 2016*, pp. 10–18, 2016.

[41] W. Cutler, "The six-piece burr," *Journal of Recreational Mathematics*, vol. 10, no. 4, pp. 241–250, 1978.

[42] D. F. Huber, "Automatic three-dimensional modeling from reality," Ph.D. dissertation, Carnegie Mellon University Pittsburgh, 2002.

[43] Q.-X. Huang, S. Flöry, N. Gelfand, M. Hofer, and H. Pottmann, "Reassembling fractured objects by geometric matching," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 569–578, 2006.

[44] H. Huang, M. Gong, D. Cohen-Or, Y. Ouyang, F. Tan, and H. Zhang, "Field-guided registration for feature-conforming shape composition," *ACM Transactions on Graphics*, vol. 31, no. 6, p. 179, 2012.

[45] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," in *Computer graphics forum*, vol. 26, no. 2, 2007, pp. 214–226.

[46] R. Toldo and A. Fusiello, "Robust multiple structures estimation with j-linkage," *ECCV 2008*, pp. 537–547.

[47] W. S. Grant, R. C. Voorhies, and L. Itti, "Finding planes in lidar point clouds for real-time registration," in *IROS 2013*, pp. 4347–4354.

[48] L. Nan and P. Wonka, "Polyfit: Polygonal surface reconstruction from point clouds," in *ICCV 2017*, pp. 2353–2361.

[49] Y. Lin, C. Wang, B. Chen, D. Zai, and J. Li, "Facet segmentation-based line segment extraction for large-scale point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 4839–4854, 2017.

[50] T. Hackel, J. D. Wegner, and K. Schindler, "Fast semantic segmentation of 3d point clouds with strongly varying density," *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 3, no. 3, 2016.

[51] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *ICRA 2011*.

[52] D. Borrmann, J. Elseberg, and A. Nüchter, "Thermal 3d mapping of building façades," in *Intelligent autonomous systems 12*. Springer, 2013, pp. 173–182.

[53] S. Choi, Q.-Y. Zhou, and V. Koltun, "Robust reconstruction of indoor scenes," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015, pp. 5556–5565.

Songlin Chen received his Ph.D. degree from the Graduate University of Chinese Academy of Sciences in 2019 and his bachelors degree from Shenyang University of Technology in 2013. He is currently an engineer at Huawei Technologies Co., Ltd. His research interests are in the fields of computer vision and deep learning.

Liangliang Nan received his Ph.D. degree from the Graduate University of Chinese Academy of Sciences in 2009 and his bachelors degree from Nanjing University of Aeronautics and Astronautics in 2003, China. He is currently an assistant professor at the Delft University of Technology (TU Delft). Prior to joining TU Delft, he was a research scientist at King Abdullah University of Science and Technology (KAUST) (2013-2017), assistant and then associate professor at SIAT (2009-2013), respectively. His research interests are in the fields of computer graphics, computer vision, and 3D Geoinformation.

Renbo Xia received his Ph.D. degree from the University of Chinese Academy of Sciences in 2006, and master degree and bachelor degree from Harbin Institute of Technology in 2002 and 2000 respectively. He is currently a research fellow at Shenyang Institute of Automation Chinese Academy of Sciences. His research interests are in the fields of computer vision and image understanding.

Jibin Zhao received his Bachelors in Mechanical Engineering at Hefei University of Technology in 1996, his Masters from Shandong University in 2000 and his Ph.D. from Graduate School of the Chinese Academy of Science in 2004, respectively. He is a researcher in Shenyang Institute of Automation, Chinese Academy of Science. His research interests include computer-aided design, rapid prototyping and revere engineering.

Peter Wonka is Full Professor in Computer Science at King Abdullah University of Science and Technology (KAUST) and Associate Director of the Visual Computing Center (VCC). Peter Wonka received his doctorate from the Technical University of Vienna in computer science. Additionally, he received a Masters of Science in Urban Planning from the same institution. After his PhD, Dr. Wonka worked as postdoctoral researcher at the Georgia Institute of Technology and as a faculty member at Arizona State University. His research interests include various topics in computer graphics, computer vision, remote sensing, image processing, visualization, machine learning, and data mining. He currently serves as Associate Editor for ACM Transactions on Graphics, IEEE Computer Graphics and Applications, and IEEE Transactions on Graphics and Visualization.